



Organizational unit: **UNFPA Evaluation Office** Year of report: **2019**

Title of evaluation report: **Evaluation of the UNFPA capacity in humanitarian action (2012-2019)**

Overall quality of report: **Very Good** Date of assessment: **2020 JUNE**

**Overall comments:** The evaluation provides a clearly presented and comprehensive assessment UNFPA's humanitarian action and capacity in a complex and evolving global system. The evaluation consistently engaged a large reference group to support the evaluation and validate findings at the global and national level, with the number of participants reaching 437 key informants in 15 countries and 150 beneficiaries across four countries in which field work was conducted. The evaluation methodology included key informant interviews, site visits and focus group discussions, and extensive document review, and is notable in its clear integration of ethical standards and GEEW considerations within the scope, methodology, findings and recommendations. Despite the limitation of access to outcome-level data, the findings were substantive and well-balanced between those that were positive and those that indicated areas for improvement in UNFPA's humanitarian action. These were translated into conclusions and recommendations that were clear and targeted. The annexes include Country Notes for each of the four countries visited as well as links to two thematic papers produced as part of the evaluation process - the six documents are all structured as evaluation reports and are well-presented (although only the main report was quality assessed). The main area for improvement would be to include more quotes and stories to highlight the voices of different groups engaged in the focus group discussions.

**Assessment Levels**

<b>Very Good</b>	strong, above average, best practice	<b>Good</b>	satisfactory, respectable	<b>Fair</b>	with some weaknesses, still acceptable	<b>Unsatisfactory</b>	weak, does not meet minimal quality standards
------------------	--------------------------------------	-------------	---------------------------	-------------	--	-----------------------	---

**Quality Assessment Criteria** *Insert assessment level followed by main comments. (use 'shading' function to give cells corresponding colour)*

<b>1. Structure and Clarity of Reporting</b>	Yes No Partial	Assessment Level:	<b>Good</b>
--	----------------------	-------------------	-------------

*To ensure the report is comprehensive and user-friendly*

<b>1. Is the report easy to read and understand (i.e. written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors?</b>	Yes	The report is very well-written and easy to understand. There are a few typographical errors in the annexed documents but these are minor and do not effect readability of the report.
<b>2. Is the report of a reasonable length? (maximum pages for the main report, excluding annexes: 60 for institutional evaluations; 70 for CPEs; 80 for thematic evaluations)</b>	Yes	The main report is 75 pages, excluding the executive summary and annexes, which falls within the maximum page requirements for thematic evaluations.
<b>3. Is the report structured in a logical way? Is there a clear distinction made between analysis/findings, conclusions, recommendations and lessons learned (where applicable)?</b>	Yes	The report is structured in a logical way with a clear distinction made between sections. While there is not a specific section on "lessons learned", the evaluation does seek to draw lessons throughout the discussion of findings and within the conclusions, and some key lessons, or notable results, drawn out in info boxes.
<b>4. Do the annexes contain – at a minimum – the ToRs; a bibliography; a list of interviewees; the evaluation matrix; methodological tools used (e.g. interview guides; focus group notes, outline of surveys) as well as information on the stakeholder consultation process?</b>	Yes	There is a separate document (Volume II) containing the annexes, and it includes all of the required elements: approach and methodology, research tools, TOR, bibliography, reconstructed theory of change, key informants interviewed, and presentation of strength of evidence of findings. The evaluation matrix as well as comprehensive and well-presented notes (which read as evaluation reports) for four countries that were part of field visits are included in the complete version of this annex. The public version of the annex does not include the matrix or country notes but a statement is provided with contact details to request these documents from the evaluation leadership. Links to the two thematic papers produced as part of the evaluation (on UNFPA supply chain management for humanitarian commodities and human resources) do not work on either version of the document but can be accessed from the UNFPA website.

*Executive summary*

<b>5. Is an executive summary included in the report, written as a stand-alone section and presenting the main results of the evaluation?</b>	Yes	The summary is clear and written as a stand-alone document with conclusions clearly linked to recommendations and providing an overview of the evaluation results.
<b>6. Is there a clear structure of the executive summary, (i.e. i) Purpose, including intended audience(s); ii) Objectives and brief description of intervention; iii) Methodology; iv) Main conclusions; v) Recommendations)?</b>	Partial	The executive summary follows the desired structure and it contains relevant information suitable for people who might not read the main report. However, the intended audience is not specified.
<b>7. Is the executive summary reasonably concise (e.g. with a maximum length of 5 pages)?</b>	Yes	The summary is close to the maximum page limit at 5.5 pages, which does not distract from the clarity and precision provided within this section.

<b>2. Design and Methodology</b>	Yes No Partial	Assessment Level:	<b>Very good</b>
----------------------------------	----------------------	-------------------	------------------

*To ensure that the evaluation is put within its context*

1. Does the evaluation describe the target audience for the evaluation?	Yes	The intended users of the evaluation are specified and primarily include stakeholders internal to UNFPA (i.e. UNFPA country and regional offices, humanitarian offices, and the Executive Board).
<b>2. Is the development and institutional context of the evaluation clearly described and constraints explained?</b>	Yes	The evaluation provides a brief overview of the institutional context and changing landscape for humanitarian work, including an analysis of the challenges this context presents. Contextual information is also clearly integrated throughout the discussion of findings.

<p><b>3. Does the evaluation report describe the reconstruction of the intervention logic and/or theory of change, and assess the adequacy of these?</b></p>	<p>Yes</p>	<p>The evaluation presented a reconstructed theory of change, as well as the process for developing it, within the annexes. The evaluation notes that there was not clear intervention logic or a theory of change on UNFPA's humanitarian action prior to the evaluation team's work.</p>
<p><i>To ensure a rigorous design and methodology</i></p>		
<p><b>4. Is the evaluation framework clearly described in the text and in the evaluation matrix? Does the evaluation matrix establish the evaluation questions, assumptions, indicators, data sources and methods for data collection?</b></p>	<p>Yes</p>	<p>The evaluation framework is briefly described in the text in terms of methodological approach and the standards and guidelines to which it conformed. It is covered in more depth within the annexes on the methodological approach (Annex I) and in the evaluation matrix (Annex V).</p>
<p>5. Are the tools for data collection described and their choice justified?</p> <p>Remember: Please address both aspects of this sub-criteria in the comment: 1) are data collection tools described (i.e. documentary review, interviews, focus group discussions etc.) and 2) is the rationale for their selection detailed</p>	<p>Yes</p>	<p>The tools for data collection and the rationale for their selection are clearly described both within the evaluation report and within Annex I. The evaluation notes that limited quantitative information on outcomes was available for the evaluation, and though an online survey was initially planned for, the constraints to implementation were significant enough to reduce the value of this tool; as such, it was removed from the mix of data collection tools implemented.</p>
<p>6. Is there a comprehensive stakeholder map? Is the stakeholder consultation process clearly described (in particular, does it include the consultation of key stakeholders on draft recommendations)?</p>	<p>Yes</p>	<p>Stakeholders were routinely consulted throughout the evaluation process. Details of the stakeholder consultation process are provided briefly within the body of the report and covered in-depth within the annexes. Flow charts are used to capture the process and timeline for stakeholder consultation and validation of findings.</p>
<p>7. Are the methods for analysis clearly described for all types of data?</p>	<p>Yes</p>	<p>The methods of analysis are described in detail in Annex I, page 12, and include descriptive analysis, content analysis, and comparative analysis.</p>
<p>8. Are methodological limitations acknowledged and their effect on the evaluation described? (Does the report discuss how any bias has been overcome?)</p>	<p>Yes</p>	<p>The methodological limitations and associated mitigation strategies are clearly described.</p>
<p>9. Is the sampling strategy described?</p>	<p>Yes</p>	<p>The sampling strategy is described in detail, and was informed by the stakeholder mapping exercise. Countries for field visits were selected purposively though in consideration of specific criteria, such as UNFPA's leadership role in the humanitarian response as well as the type and classification of the crisis in countries of operation. The evaluation was explicit in its efforts to obtain information from a representative group of stakeholders, despite the purposive nature of the sampling strategy.</p>
<p>10. Does the methodology enable the collection and analysis of disaggregated data?</p>	<p>Yes</p>	<p>The methodology enabled the collection and analysis of gender disaggregated data as focus group discussions were conducted separately with sex and age disaggregated groups. In the main report, the key informant interview stakeholders are only presented by stakeholder group and country, not by sex, and there is no indication in the KII templates that the sex of the respondent was information collected. It is noted that disaggregated participant data is available in Annex I but this is not the case. However, the annexed country report do include the breakdown by gender for both the KIIs and the FGDs.</p>
<p>11. Is the design and methodology appropriate for assessing the cross-cutting issues (equity and vulnerability, gender equality and human rights)?</p>	<p>Yes</p>	<p>The design/methodology allow for the evaluation to address cross-cutting issues, through for example, evaluation questions, regular consultation with the evaluation reference group, and through the use of multiple data collection methods (including focus group discussions with beneficiaries).</p>
<p><b>3. Reliability of Data</b></p>	<p>Yes No Partial</p>	<p>Assessment Level: <b>Very good</b></p>
<p><i>To ensure quality of data and robust data collection processes</i></p>		
<p><b>1. Did the evaluation triangulate data collected as appropriate?</b></p>	<p>Yes</p>	<p>It is noted that the evaluators triangulated data sources, data collection methods and tools, and validated data and findings through regular exchanges with the UNFPA programme staff and the Evaluation Reference Group. Document sources are regularly footnoted and qualitative findings are regularly attributed to specific stakeholder groups. Triangulation was achieved through use of multiple data sources and methods.</p>
<p><b>2. Did the evaluation clearly identify and make use of reliable qualitative and quantitative data sources?</b></p>	<p>Yes</p>	<p>Qualitative and quantitative data sources are used and consistently referenced within the report. Though the analysis heavily draws on qualitative data, this was noted as an anticipated limitation in the evaluation, and the evaluation team conducted an extensive document review to fill gaps in quantitative data, drawing on other evaluation reports conducted at the country-level. In addition, the evaluation team sought transparency by identifying the reliability and strength of evidence for each finding (Annex IV). In most cases, key informant interview data on findings was seen as reliable, however, the lack of documentation on outcome-level results severely affected the reliability of some findings as the evaluators themselves noted.</p>
<p>3. Did the evaluation make explicit any possible limitations (bias, data gaps etc.) in primary and secondary data sources and if relevant, explained what was done to minimize such issues?</p>	<p>Yes</p>	<p>The potential causes of bias are noted; the threats to reliability of the data are addressed by selecting interviewees representing a diverse range of institutional viewpoints on key topics under review. In the evaluation, 437 key informant interviews were conducted and 150 beneficiaries reached through focus group discussions.</p>
<p>4. Is there evidence that data has been collected with a sensitivity to issues of discrimination and other ethical considerations?</p>	<p>Yes</p>	<p>There is evidence that data has been collected with sensitivity to issues of discrimination. The evaluation noted adherence to UNEG and UNFPA guidelines, as well as WHO's Ethical and Safety Recommendations for Researching, Documenting and Monitoring Sexual Violence. The evaluation team also presents a table within Annex I which details how ethical and evaluation principles and standards - such as independence and impartiality, credibility and utility - were put into practice.</p>
<p><b>4. Analysis and Findings</b></p>	<p>Yes No Partial</p>	<p>Assessment Level: <b>Very good</b></p>
<p><i>To ensure sound analysis and credible findings</i></p>		
<p><b>1. Are the findings substantiated by evidence?</b></p>	<p>Yes</p>	<p>The evaluation team does a good job providing summaries of key findings for key evaluation questions in textboxes and sentences in bold at the start of relevant paragraphs. Findings are consistently substantiated by evidence, frequently citing both primary and secondary sources.</p>

2. Is the basis for interpretations carefully described?	Yes	The basis for interpretations of findings are carefully described. The findings highlight results at both the national and global levels, and notes the strength of the findings in the report and annexes. For example, the evaluation notes how the impact of UNFPA on increased awareness/knowledge of MISIP on SRHR outcomes in humanitarian settings is unknown at the global level, though some studies are highlighted at the national level which reveal results (e.g. studies conducted in DRC).
3. Is the analysis presented against the evaluation questions?	Yes	The section on findings is organized according to the evaluation questions.
4. Is the analysis transparent about the sources and quality of data?	Yes	The sources of data are consistently referenced in footnotes. Where data is missing or inadequate, the evaluation team mentions this and subsequently notes that in the conclusions and recommendations to be considered by UNFPA to ensure the quality of future evaluations and to promote learning and increased effectiveness in UNFPA's humanitarian actions.
<b>5. Are cause and effect links between an intervention and its end results explained and any unintended outcomes highlighted?</b>	Yes	The evaluators are able to show plausible cause-effect linkages and assess shortcomings, even with the constraints of outcome-level results not being systematically collected. For example, the evaluation notes how the presence of psychosocial support mobile teams are associated with increased reporting of incidences of violence and/or survivors seeking assistance. Unintended outcomes were not explicitly framed as such, but this is likely a result of the larger weakness identified of not having a clear theory of change with expected outcomes and targets.
6. Does the analysis show different outcomes for different target groups, as relevant?	Yes	The findings and analysis describe outcomes, such as service access, and ways that the activities accommodated the needs of different groups, such as youth, refugees and internally displaced persons, LGBT+ communities, and persons with disabilities.
<b>7. Is the analysis presented against contextual factors?</b>	Yes	The analysis is presented against contextual factors, as relevant.
8. Does the analysis elaborate on cross-cutting issues such as equity and vulnerability, gender equality and human rights?	Yes	As mentioned, the evaluators were careful to examine the cross-cutting issues including gender and human rights of targeted participants, particularly under the criteria of relevance and coverage.
<b>5. Conclusions</b>		
	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To assess the validity of conclusions</i>		
<b>1. Do the conclusions flow clearly from the findings?</b>	Yes	The conclusions flow clearly from the findings, with direct links made between the finding number and conclusion number.
2. Do the conclusions go beyond the findings and provide a thorough understanding of the underlying issues of the programme/initiative/system being evaluated?	Yes	The conclusions demonstrate a thorough understanding of UNFPA's humanitarian action and the Agency's position within the humanitarian arena, connecting multiple findings to one, succinct conclusion.
3. Do the conclusions appear to convey the evaluators' unbiased judgement?	Yes	There is no evidence of bias since the conclusions are clearly based on evidence from the findings showing both positive and negative elements.
<b>6. Recommendations</b>		
	Yes No Partial	Assessment Level: <b>Very good</b>
<i>To ensure the usefulness and clarity of recommendations</i>		
<b>1. Do recommendations flow logically from conclusions?</b>	Yes	The recommendations logically flow from the conclusions, and each recommendation notes the conclusions on which it is based.
<b>2. Are the recommendations clearly written, targeted at the intended users and action-oriented (with information on their human, financial and technical implications)?</b>	Yes	The recommendations are clearly written and action-oriented. The recommendations specify to whom they are directed, explicitly providing operational and technical implications.
3. Do recommendations appear balanced and impartial?	Yes	The recommendations flow clearly from the findings and conclusions without any evidence of bias.
4. Is a timeframe for implementation proposed?	Yes	Recommendations are prioritized as high, medium and low, and where possible suggest a timeframe for implementation.
5. Are the recommendations prioritized and clearly presented to facilitate appropriate management response and follow up on each specific recommendation?	Yes	Recommendations are clearly prioritized and phrased so as to facilitate management response.
<b>7. Gender</b>		
	0 1 2 3 <sup>(95%)</sup>	Assessment Level: <b>Very good</b>
<i>To assess the integration of Gender Equality and Empowerment of Women (GEEW) <sup>(*)</sup></i>		

<p>1. Is GEEW integrated in the evaluation scope of analysis and indicators designed in a way that ensures GEEW-related data to be collected?</p>	<p><b>a. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives? (Score: 0-3)</b> The assessment of gender and human rights was included as a specific objective of the evaluation. (Score = 3).</p> <p><b>b. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria? (Score: 0-3)</b> While there was no standalone criterion for the assessment of gender and/or human rights, it was mainstreamed into the criteria of relevance and coverage. (Score = 3).</p> <p><b>c. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation? (Score: 0-3)</b> Several evaluation questions explicitly integrate GEEW. (Score = 3)</p> <p><b>d. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?(Score: 0-3)</b> An assessment of the sufficiency of information collected during the implementation period is addressed, including whether needs assessments were conducted and used in the design of interventions. Gaps were noted in this regard. (Score = 3).</p>
<p>2. Is a gender-responsive methodology used, including gender-responsive methods and tools, and data analysis techniques?</p>	<p><b>a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex? (Score: 0-3)</b> The evaluation methodology was explicit in how gender and human rights were incorporated into the methodology and analysis methods to ensure the collection of disaggregated data. However, the overall number of evaluation participants is not sex disaggregated (this information was included for each country report) (Score = 2).</p> <p><b>b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEEW considerations (collecting and analyzing both quantitative and qualitative data, and ensuring the appropriate sample size)? (Score: 0-3)</b> The methods and data collection tools included key informant and focus group interviews with stakeholders and primary beneficiaries that included women, youth and refugees, with males and females interviewed as separate groups. The sample size was appropriate for an evaluation with a global scope. (Score = 3).</p> <p><b>c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility? (Score: 0-3)</b> Diverse data sources and multiple methods are used, including focus groups, observations, document review, and key informant interviews. Data is triangulated and validated by a larger and representative evaluation reference group. (Score = 3).</p> <p><b>d. Do the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate? (Score: 0-3)</b> Perspectives of users of interventions, including diverse vulnerable groups, were obtained through FGDs. (Score = 3).</p> <p><b>e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality? (Score: 0-3)</b> The interview and focus groups protocols indicate that data collectors complied with ethical guidelines for informed consent and respect for confidentiality. Ethical standards, including WHO guidelines for researching sexual violence, were integrated into the evaluation methodology. (Score = 3).</p>
<p>3. Do the evaluation findings, conclusions and recommendations reflect a gender analysis?</p>	<p><b>a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality? (Score: 0-3)</b> The subsection on the background of humanitarian action does not specifically articulate how some groups are more marginalized, though this is taken up in the findings. (Score = 2).</p> <p><b>b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable? (Score: 0-3)</b> The perspectives of different groups are routinely and transparently referenced, though there is limited reference to quotes and stories drawn from the focus group discussions in the main report or the county reports. (Score = 2)</p> <p><b>c. Are unanticipated effects of the intervention on human rights and gender equality described? (Score: 0-3)</b> The evaluators did not specifically identify any findings as being unanticipated, though this is likely a result of the absence of a clear theory of change. (Score = 2).</p> <p><b>d. Does the evaluation report provide specific recommendations addressing GEEW issues, and priorities for action to improve GEEW or the intervention or future initiatives in this area? (Score: 0-3)</b> Recommendation 6 specifically addresses GEEW and human rights issues and priorities for action. (Score = 3).</p>

(\*) This assessment criteria is fully based on the UN-SWAP Scoring Tool. Each sub-criteria shall be equally weighted (in correlation with the calculation in the tool and totaling the scores 11-12 = very good, 8-10 = good, 4-7 = Fair, 0-3=unsatisfactory).

(\*\*) Scoring uses a four point scale (0-3).  
0 = Not at all integrated. Applies when none of the elements under a criterion are met.  
1 = Partially integrated. Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.  
2 = Satisfactorily integrated. Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.  
3 = Fully integrated. Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

**Overall Evaluation Quality Assessment**

Quality assessment criteria (scoring points*)	Assessment Levels (**)			
	Very good	Good	Fair	Unsatisfactory
1. Structure and clarity of reporting, including executive summary (7)		7		
2. Design and methodology (13)	13			
3. Reliability of data (11)	11			
4. Analysis and findings (40)	40			
5. Conclusions (11)	11			

6. Recommendations (11)	11			
7. Integration of gender (7)	7			
<b>Total scoring points</b>	<b>93</b>	<b>7</b>		
<b>Overall assessment level of evaluation report</b>	<b>Very Good</b>			
	<b>Very good</b> very confident to use	<b>Good</b> confident to use	<b>Fair</b> use with caution	<b>Unsatisfactory</b> not confident to use

(\*) (a) Insert scoring points associated with criteria in corresponding column (e.g. - if 'Analysis and findings' has been assessed as 'Good', enter 40 into 'Good' column.  
(b) Assessment level with highest 'total scoring points' determines 'Overall assessment level of evaluation report'. Write corresponding assessment level in cell (e.g. 'Fair').  
(c) Use 'shading' function to give cells corresponding colour.

If the overall assessment is 'Fair', please explain

• How it can be used?

• What aspects to be cautious about?

**Where relevant, please explain the overall assessment Very good, Good or Unsatisfactory**

The evaluation was well done without any significant omissions.

**Consideration of significant constraints**

The quality of this evaluation report has been hampered by exceptionally difficult circumstances:

Yes       No

If yes, please explain: